Information Superiority through Data Warehousing

Neil Warner
ADI Limited
Neil.Warner@adi-limited.com

Abstract

The aim of a Command Support System is to achieve decision superiority by providing a knowledge edge to the Commander. A precursor to a knowledge edge is Information Superiority.

Within most current Command Support Systems minimal integration and fusion of data is undertaken to provide the basis of information superiority. Most data remains in its island and any integration of data is provided by human intervention or through the use of specific and often limited applications.

Despite the introduction of computer systems and communication networks, decision makers cannot readily access critical information that already exists in an organization. Data is locked up in a myriad of computer systems and is exceedingly difficult to get at - "data in jail". Only a small fraction of the data that is captured, processed and stored in the enterprise is actually available to decision makers.

Data Warehousing provides technology that has evolved into a new technology making it possible to attack the problem of providing commanders with access to whatever level of information is needed to make decisions.

This paper will address the concept of providing improved information superiority and therefore decision superiority by the use of the commercial information technology “data warehousing” concepts and tools.

Introduction

The aim of a Command and Control System and Command Control System (generically C4I Systems) is to provide decision superiority by providing a knowledge edge to the Commander. A knowledge edge is provided by a superior and effective combination of People, Doctrine and Technology. A precursor to a knowledge edge is Information Superiority. This is diagrammatically represented, in Australian Defence Organization terms, in Figure 1 below.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2001		**00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Information Superiority through Data Warehousing	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **ADI Limited,1133 Regent Street Suite 300,Fredericton, NB E3B 3Z2 Canada, ,**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**14**	

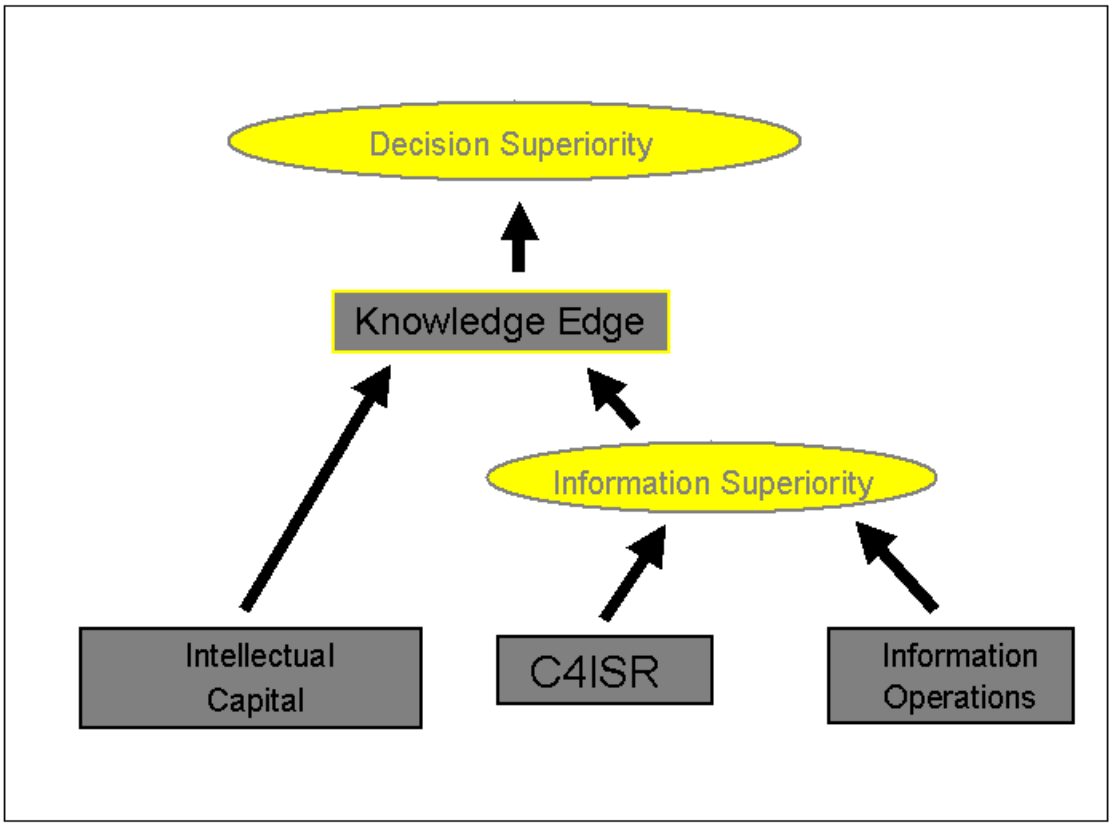

Figure 1 – Information Superiority[1]

Within most current Command Support Systems (CSS) minimal integration and fusion of data is undertaken or provided to produce the basis of information superiority. Most data remains in its island and integration of data is only provided in some pre-defined applications or by human intervention. Data is not distributed throughout the system and data can be restricted to one node or Headquarters. It is difficult for new users and some experienced users to find data within the system. In general, C4I Systems do not currently provide a robust basis for obtaining information superiority or knowledge edge.

To be effective in the future, it will be necessary for Commanders and their staff to analyse, plan and react to changing conditions in a much more rapid manner. Despite the introduction and use of powerful computer systems and communication networks that span the globe, decision makers cannot get their hands on critical information that already exists in an organization. Data is locked up in a myriad of computer systems and is exceedingly difficult to get at - "data in jail". Only a small fraction of the data that is captured, processed and stored in the enterprise is actually available to decision makers.

[1] [Nicholson, 2000]

This can be overcome by the use of data warehousing to fuse both operational and informational data. An important aspect of achieving this is through the architecture of the data warehouse that facilitates the required access. However, the manner in which the information is exploited is also important in that this reflects to a large extent the types of capabilities that need to be achieved if information superiority is to be achieved.

Aim

This paper will address the concept of providing improved information superiority and therefore decision superiority by the use of the commercial information technology "data warehousing" concepts and tools.

Definitions

Information Superiority: A relative state achieved when a competitive advantage is derived from the ability to exploit an information advantage. The ability to develop and use information while denying an adversary the same capability

Information Fusion: Events, activities and movements will be correlated and analyzed as they occur in time and space, to determine the location, identity and status of individual objects (equipment and units), to assess the situation, to assist in threat assessment and to detect patterns in activity that reveal intent or capability.

Information Types

Most information systems[2] handle two different types of data within the system, "Operational Data" and "Informational Data" Both are important in decision superiority, but there applicability to Data Warehousing and related techniques requires an examination and understanding of their characteristics in terms of the structure and nature of the associated data as they refer to C4I systems.

Most Command Systems provide "Operational" data as opposed to "Informational" data. Both these types of data are important in the provision of Information Superiority. Operational Data aids in day-to-day activities. Examples of operational data would be:

a) Track, contact or positional data;
b) Logistic data detailing stores holdings or supply provisioning data.

[2] [Orr, 1997]

Informational Data is mainly used in making decisions and is often consists of reference material or historical data. Examples of Informational Data are:

a) Janes Fighting Ships;
b) CNN News Broadcast summaries of last two months.

Operational and Informational Data has different focus and scope. Informational data requires historical reference to be of full use. Where operational data needs are normally focused upon a single area, informational data needs often span a number of different areas and need large amounts of related operational data.

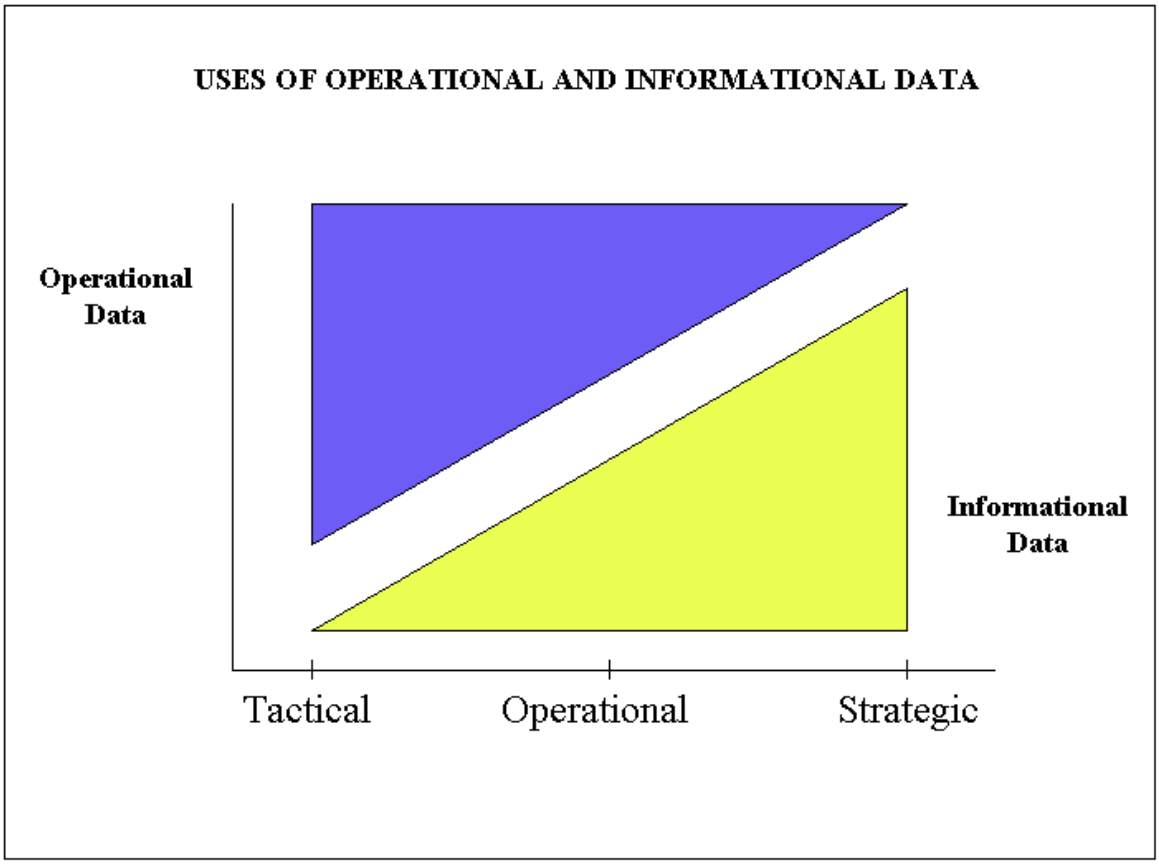

Figure 2 – Data Type at Different Levels of Command

Data Warehousing

Data Access Crisis[3]

[3][Orr, 1997]

If there is a single key to survival in the 1990s and beyond, it is being able to analyze, plan and react to changing operational and business conditions in a much more rapid fashion. To do this, top managers, analysts and knowledge workers in our enterprises need more and better information. Information technology itself has made possible revolutions in the way that organizations today operate throughout the world. But the sad truth is that in many organizations despite the availability of more and more powerful computers on everyone's desks and communication networks that span the globe, large numbers of executives and decision makers cannot get their hands on critical information that already exists in an organization. Every day organizations large and small create billions of bytes of data about all aspects of their business, millions of individual facts about their customers, products, operations and people. But for the most part, the data are locked up in a myriad of computer systems and are exceedingly difficult to access. This phenomenon has been described as "data in jail". Experts have estimated that only a small fraction of the data that is captured, processed and stored in the enterprise is actually available to executives and decision makers. While technologies for the manipulation and presentation of data has literally exploded, it is only recently that those involved in developing IT strategies for large enterprises have concluded that large segments of an enterprise are often "data poor."

Providing Data Access to the Enterprise

Recently, a set of significant new concepts and tools has evolved into a new technology that makes it possible to attack the problem of providing all the key people within an enterprise with access to whatever level of information is needed for the enterprise to survive and prosper in an increasingly competitive world. These experiences have allowed the IT industry to identify what are the key problems that have to be solved.

The term that has come to characterize this new technology is "data warehousing". Data warehousing has grown out of repeated attempts on the part of various researchers and organizations to provide flexible, effective and efficient means of accessing the sets of data that have come to represent one of an organizations most critical and valuable assets.

The results is a series of concepts, technologies and tools that have be integrated to form a system that provides the desired functionality and process. This is illustrated in Figure 3.

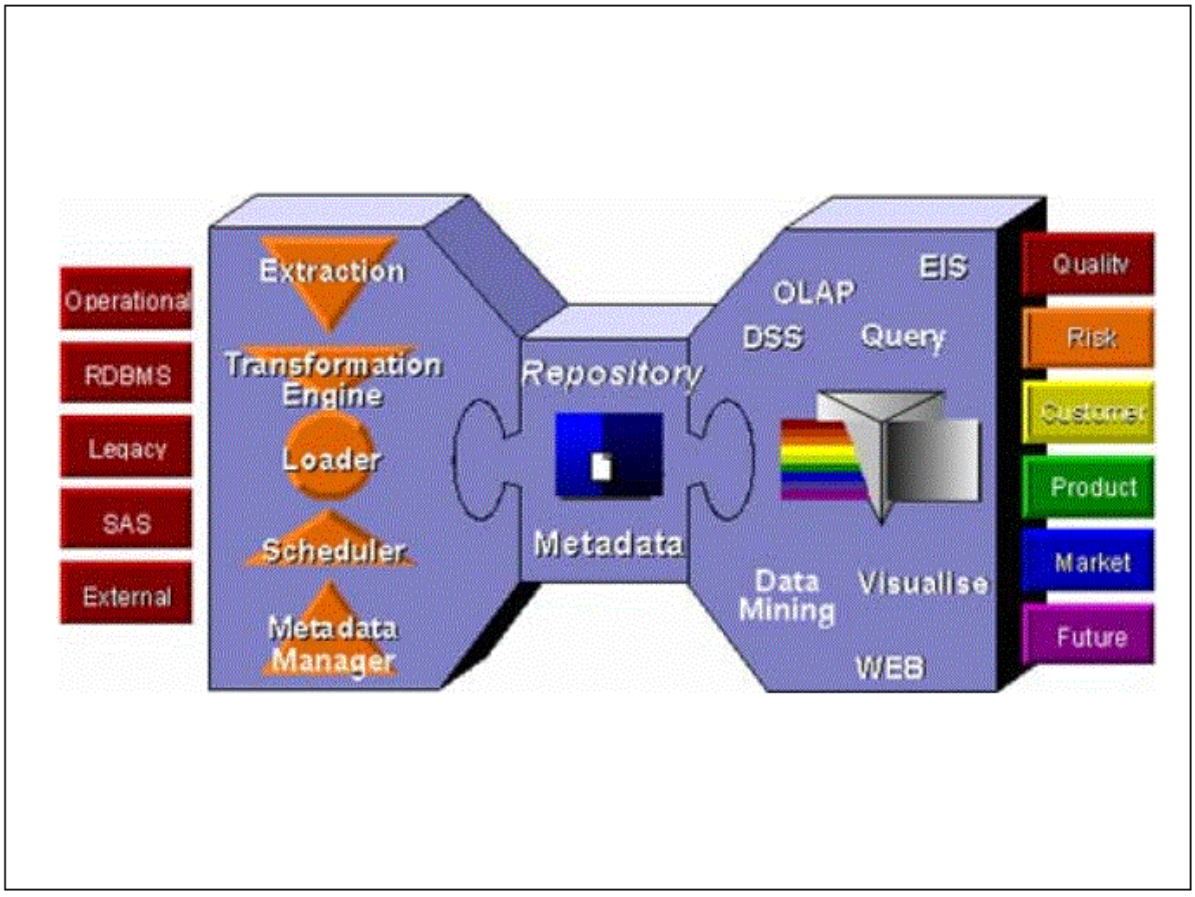

Figure 3 – Data Warehousing Technology Concepts[4]

Architecture of a Data Warehouse in a C4I Environment

General

One unfortunate consequence of data warehousing is the name with its emphasis on on "data" rather than "information" This leads to the assumption that data warehousing can be considered as a database issue, which it is not. Data warehousing must be considered as a technique that allows useful relationships to be established between operational data and informational data and this can only be achieved by understanding the architectural requirements.

Factors Affecting the Architecture

The following factors should be considered when examining the architecture for data Warehousing within a C4I environment:

[4] [SAS Institute, 1999]

a) Operational Concepts, including Operating Procedures of Headquarters and the Strategic/Operational/Tactical Mix of HQ's;
b) Network Topology, including both Bandwidth and Reliability considerations;
c) Redundancy Requirements;
d) Ease of Access;
e) Speed of Access;
f) Security - Need to Know.
g) Precedence of access within the user community

In general we tend to think in terms of three broad categories of end-users of Data Warehouses:

a) Commanders and Senior Staff Officers
b) "Power" users (Analysts and Planners etc.)
c) Support users (clerical, administrative, etc.)

Each of these different categories of user has its own set of requirements for data, access, flexibility and ease of use and therefore will affect the architecture of the eventual data warehouse.

Architectural Overview

One possible architecture for data warehousing is detailed in Figure 4. Although this only gives a very high level overview, it does provide a reference for discussion in the context of this paper. The key features of this architecture are:

a) The user is presented with common user interface regardless of data being accessed, probably a Web Browser;
b) User access is through the distributed Metadata Repository which could be distributed to multiple sites - Metadata Repository could provide some level of data aggregation;
c) Data is warehoused based on a common schema as well as user defined schema and profiles, so that not all data is warehoused, but only data indicated as significant;
d) One Central Repository for actual data, which may be composed on several different data stores:
    i) Structured Data – Oracle,
    ii) Semi Structured Data - Lotus Notes,
    iii) Unstructured Data – Web;
e) Data compression where appropriate is used in Data Stores, although this is an implementation issue.

This is, however, only one possible architecture and any decision on the most appropriate architecture for a data warehouse is dependent upon the overall management approach, system architecture and user requirements.

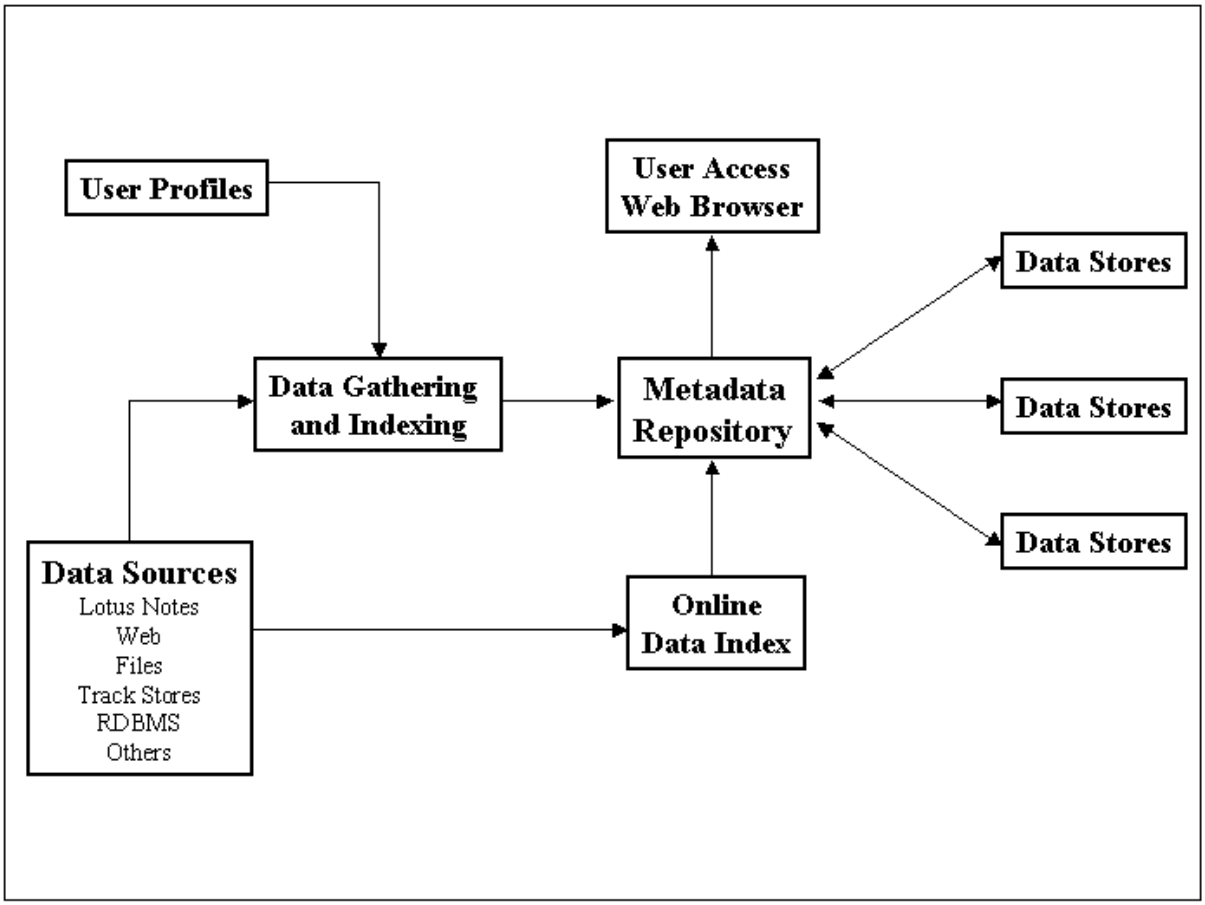

Figure 4 – Possible Architecture – C4I Data Warehousing

Management

Until relatively recently there were those that believed information management could be achieved for decision support by pointing users directly at the operational data (even though inevitably the actual information was not there or was in the wrong form) or that merely copying operational transaction data and storing it over time would be sufficient. This resulted in extremely large (therefore slow) data dumps rather than useful information repositories.

Most importantly, the data warehouse management should also support an open metadata layer as its foundation. Metadata is effectively - data about data as well as information and metrics. It enables the users of the data warehouse to understand what they have access to, where it comes from, when it is updated, who owns it, etc.

Once established, metadata becomes an organizational asset where business rules and definitions are stored.

Information Fusion and Exploitation

It is important to realize that data warehousing does not provide by itself information superiority. But there is a lot of implied, or buried information in a data warehouse that can be used to obtain information superiority. Often it is a question of using information fusion and the appropriate exploitation tools to find the information that is required.

Data Exploitation

Data exploitation can be regarded as falling into two broad, overlapping categories, reporting and applied analysis.

For reporting, this can be general querying of the information repository and generating a report, to more sophisticated multidimensional viewing, which is often included in such applications as BI (Business Intelligence), EIS (Executive Information Systems) or even OLAP (On-Line Analytical Processing with multidimensional viewing).

For example, the graph detailed in Figure 5 was created using Excel querying a data warehouse, and shows a count of track types within a particular geographic area against time. This example shows that the exploitation of the data warehouse can be used to view, understand and use available information and make comparisons to previous periods.

The other broad category of applied analysis is becoming popularized under the heading "data mining". As with many other aspects of data warehousing, the use of sophisticated statistical analysis to uncover relationships and trends within data is not new. However, combining a complete suite of tools within a user (analyst) friendly graphical environment for the complete process of data mining is.

An example of applied analysis is shown in Figure 6. In this example, the anlysis is attempting, using some data assumptions, whether any further data fusion and correlations can be made. The data involved is again track position and time data. Track correlation can be easily detected from the resulting 3 dimensional graph, as the diagonal line within the graph. Other less obvious correlations can also be made. The use of colours to display data grouping assumptions assists in the visual detection of result features.

Solutions are coming onto the market that guide the analyst from sampling, exploring, manipulating data, suggesting the best model for assessing the results of analysis. This is an exciting development that adds great value to data warehousing by uncovering relationships in information that help us understand our business and customers better.

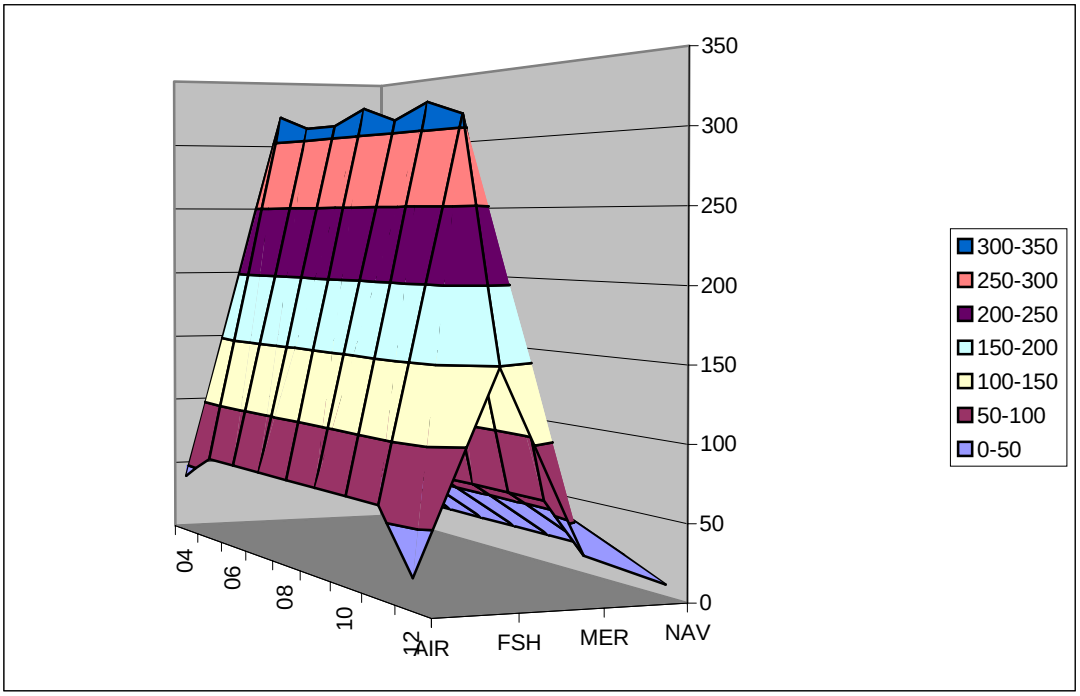

Figure 5 – Example of Simple Reporting Category of Exploitation

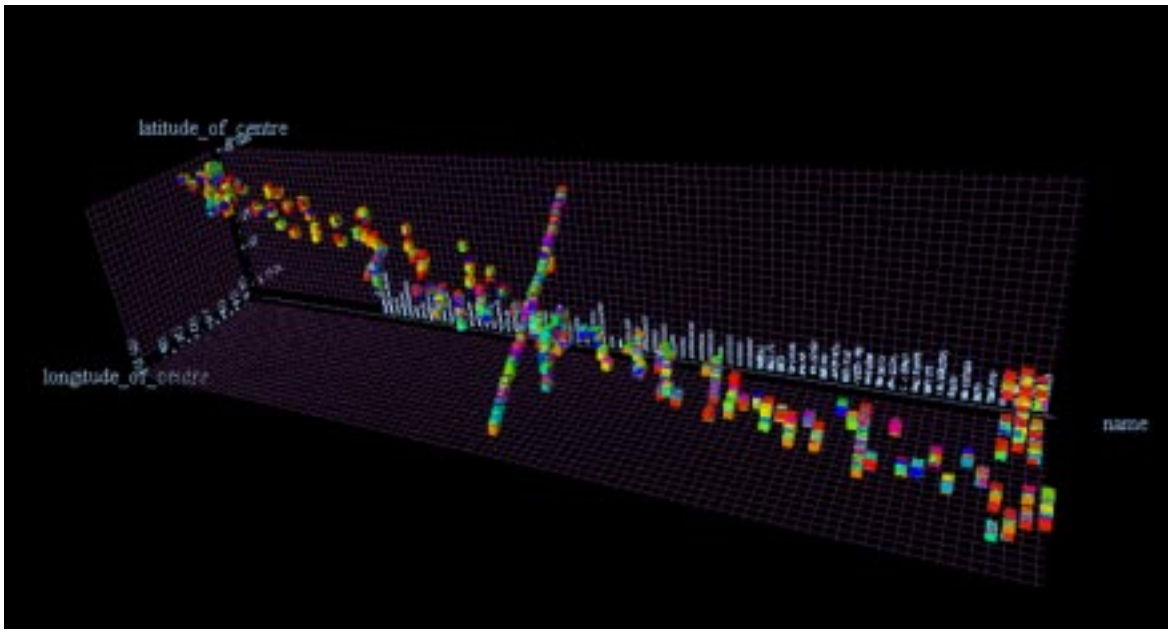

Figure 6 – Example of Simple Applied Analysis Category of Exploitation

Information and Data Fusion

A data warehouse will be of little benefit to the warfighter unless some assistance is given to be able to access the data provided in an intelligent way.

The basic requirements of this capability are:

a) The capability to fuse both operational and informational data;
b) Showing the relationship and the context between operational and informational data;
c) The system will need to be able to organize data to create information and be able to present it to the users in a manner that will allow easy access to the information;
d) Automatic discovery of the content within an organization or HQ by mining it, analyzing its meaning, and organizing it into a meaningful taxonomy of related categories.

Knowledge Management concepts are also being tried as another means of providing information fusion. Specific technologies will also be required to refine, direct and manage the information fusion capabilities. At present these technologies are at best immature, although indications are that commercial companies are working to this requirement and COTS Software should be released shortly that may go some way to solving the information fusion problem.

Perceived Operational Benefits

Access to more data or information does not in itself, provide any operational benefit to the warfighter. Information superiority would only be provided if the data and information is presented to the warfighter in a manner that makes it easier to carry out effectively the required function or operation. The key requirements are the manner in which the information is presented, as well as access to the required information, both current and historically based data. A distributed meta data repository is suggested as one of the more effective means of meeting the two key requirements.

The provision of historical data to the warfighter is an important issue. This can give the user a historical perspective of an event, but this can also be related to the action undertaken previously to a given set of events. The provision of historical data can provide the facilities to undertake trend analysis, not only for intelligence analysis but also for operational analysis for post operation or exercise analysis. This can also provide the basis of a “lessons learnt” repository for future training and evaluation.

Effects on Information Architecture

The implementation of data warehousing within a C4I system will require modification to the information architecture employed within the C4I Systems.

Elements of the data warehouse structure could be used to access data stored within the C4I system, but not in the data warehouse. A portal like mechanism could be used to provide a window to current data within the system. Within the architecture shown at Figure 4, the "Online data Index" Black Box could be implemented with Web Based Portal Technology. The use of a data warehouse could also move for the integration of different data stores within the system.

The use of a data warehouse within the information architecture of a C4I system can be implemented in a flexible manner depending on the requirements of the C4I system. The key to this flexibility is the use of a meta data repository. Within the proposed architecture, the ability to access information outside the data warehouse does much to increase data accessibility and overcome data in jail problems identified earlier. Perhaps, more importantly, the flexibility in the use of a data warehouse underlines the need for architects of C4I systems to understand that any architecture must be dynamic and dictated by daily use, not by a prior concepts about the possible nature of operations.

Outstanding Issues

Data warehousing is a new concept and it is too early to make broad judgements on its efficacy. This must wait until tools and techniques such as those described in this paper are more fully developed and used. What it does attempt to do is overcome the vexed problem of providing the war fighter with more data, but not necessarily more information. The combination of access to on-line data and data within a data warehouse does, however, mean that we are beginning to be able to view events as they occur within a C4I system in context. No longer is the analysis fragmented, the capability is emerging for the war fighter to manage the analysis in terms of actual requirements. The challenge now, is to ensure that this leads to better decision making and thereby information superiority that can be claimed with confidence.

Conclusion

Data warehousing provides concepts and tools that have evolved into a new technology making it possible to attack the problem of providing all the key commanders within the enterprise with access to whatever level of information is needed to make robust decisions. Data warehousing has evolved through repeated attempts on the part of various

researchers and organizations to provide their organizations flexible, effective and efficient means of getting at the sets of data that have come to represent one of the organizations most critical and valuable assets. Data warehousing is a technology that has grown out of the integration of a number of different technologies and experiences over the last two decades. Data warehousing technology can provide an Enterprise Wide Framework for managing data within a Command and Control Organization.

Data warehousing does not provide information superiority or the knowledge edge on its own. It has the potential to give the C4I user access to the buried and implied knowledge that can be held within the data warehouse. It is a question of accessing the information within the data warehouse through On-Line Analytical Processing (OLAP) and data mining technology with decision support tools that will provide the advantage to the commander. The information within the data warehouse may need to be transformed through different views to reflect the real dimensionality of the situation as understood by the commander.

Data Warehouse does not provide all the answers for information superiority in a C4I System, it does, however, provide a sound basis for obtaining it. Intelligent Tools and Knowledge Techniques are required to fully exploit data held within the Warehouse.

References

[Nicholson, 2000] AVM Nicholson RAAF. "The Defence Information Environment". Australian C3I Research Forum, August 2000.

[Orr, 1997], Ken Orr, "Data Warehousing Technology, a White Paper", The Ken Orr Institute, 1997, http://www.kenorrinst.com/dwpaper.html.

[Boar, 1999], Bernard Boar, "Understanding Data Warehousing Strategically", http://www3.ncr.com/product/retail/presentations/whitepapers/data_warehousing/index.shtml

[Burlton, 1999], Roger Burlton, "Data Warehousing in the Knowledge Management Cycle',
http://datawarehouse.dci.com/articles/1998/08/25km2.htm

[SAS Institute, 1999], SAS Institute, "Data Warehousing Fundamentals",
http://technology.news.com.au/concepts/warehou.htm